

Notice of Allowability	Application No.	Applicant(s)	
	10/780,235	FORMAN ET AL.	
	Examiner Cheryl Lewis	Art Unit 2167	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTO-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. This communication is responsive to the applicants communication filed on November 13, 2006 and January 3, 2007.
2. The allowed claim(s) is/are 1-7, 9-11, 13-19, and 22-25, renumbered as claims 1-21.
3. Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 - a) All
 - b) Some*
 - c) None
 of the:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

* Certified copies not received: _____.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

4. A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.
5. CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
 - (a) including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached
 - 1) hereto or 2) to Paper No./Mail Date _____.
 - (b) including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____.

Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).
6. DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. Notice of References Cited (PTO-892)
2. Notice of Draftsperson's Patent Drawing Review (PTO-948)
3. Information Disclosure Statements (PTO/SB/08),
Paper No./Mail Date _____
4. Examiner's Comment Regarding Requirement for Deposit
of Biological Material
5. Notice of Informal Patent Application
6. Interview Summary (PTO-413),
Paper No./Mail Date 1/15/07.
7. Examiner's Amendment/Comment
8. Examiner's Statement of Reasons for Allowance
9. Other _____.

DETAILED ACTION

1. Claims 1-7, 9-11, 13-19, and 22-25 are allowed. These claims have been renumbered as claims 1-21.
2. The applicants have cancelled claims 12, 20, and 21 in the amendment received on November 13, 2006.

Drawings

3. The drawings filed on February 17, 2004 are accepted by the Examiner.

EXAMINER'S AMENDMENT

4. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in an interview with Dan Hu on January 16, 2007.

5. Claims 1, 9, 11, 17, and 19 have been amended and claim 8 has been cancelled as follows:

1. (Currently Amended) A processor-implemented method for generating masks for data de-duplication from entity eonym data fields in a given set of data

records, said data records each including an entity location data field, the method comprising:

for each data record, splitting each entity eponym data field into a corresponding prefix-suffix combination, and for each prefix, a processor computing a tally of distinct entity locations, and for each prefix and entity location combination, the processor computing a tally of distinct suffixes; and

setting, by the processor, a threshold boundary wherein a prefix is defined as one of said masks when one or more of the tallies are indicative of different eonyms signifying a particular entity, wherein the one mask enables a particular data record to be matched to the particular entity by ignoring a portion of the particular data record, wherein said de-duplication involves matching each data record representing a specific activity to the particular entity of a plurality of known entities such that duplication of entities is reduced in a database of said plurality of known entities.

8. (Cancelled)

9. (Currently Amended) The method as set forth in claim 8 1 wherein said masks are generated as rules for ignoring variable data portions of the entity eonym data fields and assigning a respective data record therefor to said database based on a non-variable data portion of the corresponding entity eonym data field.

11. (Currently Amended) A processor-implemented method for partitioning a plurality of data packets in a database such that duplication of data groups is minimized, the method comprising:

selecting a primary identifier data field and a secondary identifier data field for each data packet that represents a corresponding activity;

for all data packets having a non-unique primary identifier data field, using heuristic procedures for splitting each primary identifier data into at least one prefix-suffix combination;

for each prefix, counting a first tally of how many distinct secondary identifier data fields occurs, and counting a second tally of how many distinct secondary identifier data fields occur with a single suffix, and for each prefix and each secondary identifier data field matched thereto, counting a third tally of how many distinct suffixes occur;

based on said first tally, said second tally and said third tally generating masks representative of prefixes applicable to said data packets having a non-unique primary identifier data field such that application of said masks assigns data packets having a non-unique primary identifier data field to associated common entities defined thereby, wherein application of said masks provides cleaning of the data packets; and

filling each of said data packets into a single file assigned to respective said associated common entities defined.

17. (Currently Amended) A processor-implemented method of doing business of data de-duplication comprising:

receiving, by a processor, a periodic log of transactions representing corresponding activities associated with entities, each transaction represented by a data string including at least a name field and another identifier field;

selecting, by the processor, unique representative samples of said transactions; for each of said samples, the processor dissecting each name field into a corresponding prefix and suffix combination, and for each prefix and each another identifier combination, the processor counting a number of distinct suffixes and storing a tally therefor; and

generating, by the processor, a mask from a specific prefix when the specific prefix meets a predefined decision criteria which is a function of said tally, wherein the mask is applicable to the log of transactions to enable at least some of the data strings to be matched to a particular entity name by ignoring variable portions of the at least some data strings such that duplication of entities is reduced.

19. (Currently Amended) A computer memory containing instructions that when executed cause a computer to:

store a given set of data records representing activities for a given set of entities, each of said data records having discrete data fields including an entity identification field and an entity location field;

split each entity identification field into a corresponding prefix-suffix combination; for each prefix, compute a tally of distinct entity locations; for each prefix and entity location field combination, compute a tally of distinct suffixes therefor;

set a threshold boundary wherein a prefix is defined as one of said masks when one or more of the tallies is indicative of different entity identification strings in entity identification fields signifying a single one of said entities; and

apply said masks to said given set of data records such that each record is assigned to a corresponding one of said given entities, wherein applying the masks provides cleaning of the data records.

REASONS FOR ALLOWANCE

6. The following is a statement of reasons for the indication of allowable subject matter:

Applicants' response filed on November 13, 2006 overcomes the prior art rejection under 35 USC § 102(b) by Walker et al.

The prior art of record does not render obvious to one ordinarily skilled in the art at the time of applicant's invention nor anticipate the combination of claimed elements including 'wherein said de-duplication involves matching each data record representing a specific activity to the particular entity of a plurality of known entities such that duplication of entities is reduced in a database of said plurality of known entities' as recited in independent claim 1.

As per claim 11, the prior art of record does not teach 'based on said first tally, said second tally and said third tally generating masks representative of prefixes applicable to said data packets having a non-unique primary identifier data field such that application of said masks assigns data packets having a non-unique primary identifier data field to associated common entities defined thereby, wherein application of said masks provides cleaning of the data packets'.

As per claim 17, the prior art of record does not teach 'generating, by the processor, a mask from a specific prefix when the specific prefix meets a predefined decision criteria which is a function of said tally, wherein the mask is applicable to the log of transactions to enable at least some of the data strings to be matched to a particular entity name by ignoring variable portions of the at least some data strings such that duplication of entities is reduced'.

As per claim 19, the prior art of record does not teach 'apply said masks to said given set of data records such that each record is assigned to a corresponding one of said given entities, wherein applying the masks provides cleaning of the data records'.

The remaining claims, 2-7, 9, 10, 13-16, 18, and 22-25 are dependent claims, thus these claims are patently distinct over the art of record for at least the above reasons.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

NAME OF CONTACT

7. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Cheryl Lewis whose telephone number is (571) 272-4113. The examiner can normally be reached on 6:30-3:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Cottingham can be reached on (571) 272-7079. The fax phone number for the organization where this application or proceeding is assigned is (571) 273-8300.

(571) 273-4113 (Use this FAX #, only after approval by Examiner, for "INFORMAL" or "DRAFT" communication. Examiners may request that a formal paper/amendment be faxed directly to them on occasions.).

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist/ Technology Center (571) 272-2100.

Art Unit: 2167

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see <http://pair-direct.uspto.gov>.

Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).



Cheryl Lewis
Patent Examiner
January 16, 2007